Group Sparsity in SAR Tomography – Experiments on TanDEM-X Data Stacks

Xiao Xiang Zhu[1,2], Nan Ge [1], Muhammad Shahzad [2]

(1) Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, D-82234 Wessling, Germany (Email: xiao.zhu@dlr.de)

(2) Helmholtz Young Investigators Group: SiPEO, Chair of Remote Sensing Technology, Technische Universität München

Abstract: *With meter-resolution images delivered by modern SAR satellites like TerraSAR-X and TanDEM-X, it is now possible to map urban areas from space in very high level of detail using advanced interferometric techniques such as PSI and (Tomographic SAR) TomoSAR inversion, whereas these multi-pass interferometric techniques are based on a great number of images. We are aiming at improving the estimation accuracy of TomoSAR while reducing the required number of images. In the paper, we propose a novel workflow that marries the freely available 2D building footprint GIS data and the group sparsity concept for TomoSAR inversion. Experiments on bistatic TanDEM-X data stacks demonstrate the great potential of the proposed approach.*

1. Introduction

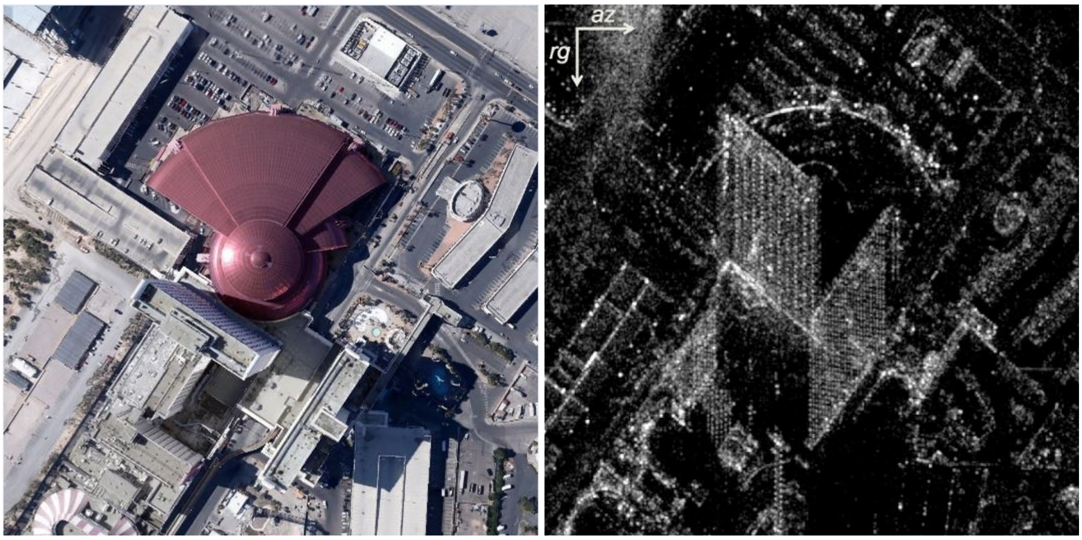
Modern spaceborne synthetic aperture radar (SAR) sensors, such as TerraSAR-X, TanDEM-X and Cosmo-Skymed, deliver SAR data with very high spatial resolution of up to 1 m compared to the medium (10–30 m)- and high (3–10 m)-resolution SAR systems available so far. With these meter-resolution data, advanced multi-pass interferometric techniques such as persistent scatterer interferometry (PSI) and tomographic SAR (TomoSAR) inversion allow retrieving not only the 3D geometrical shape but also the undergoing temporal motion in the scale of millimeter of individual buildings [1]–[4]. In particular, sparse reconstruction based methods [5] [6], like SL1MMER, are characterized by robust TomoSAR inversion with very high elevation resolution, and can offer so far ultimate 3D and even 4D SAR imaging [7].

The downside of PSI and TomoSAR is their high demand on data, i.e., typically a stack of 20–100 images over the illuminated area are required. However, if we can extract certain detailed features or patterns of the high rise buildings in the SAR images, the required number of images can be significantly reduced by incorporating such features as prior in the estimation.

For this purpose, we propose a novel workflow marrying the globally available 2D building footprint GIS data and the group sparsity concept for TomoSAR inversion. In the first stage, online freely assessable 2D building footprints are used for extracting detailed high rise building features including building masks, orientations, and the iso-height lines in SAR image stacks. Then, the group sparsity model, named as M-SL1MMER, is employed for joint TomoSAR inversion of the identified iso-height pixels. The proposed approach is validated using TanDEM-X data stacks. Compared to the single-snapshot sparsity model, as used in SL1MMER, the superior performance of the proposed M-SL1MMER approach in terms of super-resolution power and robustness are evident.

2. Data Set

We work with 11 bistatic interferograms acquired by TerraSAR-X and TanDEM-X, with cross-track baselines ranging from approximately -200 to +200 m. The single-pass characteristic renders path delay effects very small and deformation negligible. Figure 1 shows an optical image of the test area together with the corresponding SAR mean intensity image.



(a) (b)

Figure 1. Test Area: (a) optical image of the test area © Google; (b) corresponding SAR intensity map (rg and az refer to range and azimuth coordinates, respectively).

3. Prior Knowledge Retrieval

In order to retrieve prior information pertaining to building regions, 2D building footprints are downloaded from OpenStreetMap (OSM). OSM is a free and open project containing global GIS data with topological data structure. Building footprints are represented as polygons with ordered list of nodes/vertices (i.e., pairs of UTM or latitude/longitude coordinates according to WGS 84 coordinate system).

3.1 Extraction of Building Mask in SAR image

The key idea is to incorporate these available building footprints as prior knowledge into the TomoSAR inversion such that higher accuracy can be achieved even with reduced the number of images. In this regard, the first necessary step is to perform 2D transformation to project the vertices of these footprints from world coordinates to SAR range/azimuth coordinates. Figure 2 shows the result of the reference polygon overlaid onto the building of interest in the corresponding SAR image of Figure 1. Due to strong layover effect, as evident from the Figure 2, the projected building footprint cannot be directly used for further processing and therefore overlaid pixels belonging to the same building needs to be segmented. To achieve this, following recursive procedure is adopted:

1) Firstly the side of the building footprint facing the SAR sensor is identified. If we assume that $v_{i=1,\ldots,n}$ denote the indices of ordered 2D footprint vertices of one particular building. Then any vertex v_k $(k \in n)$ belongs to the side facing the sensor if and only if its projection onto the line at zeroth range axis (i.e., line defined as $rg = 0$ with zero azimuth slope) does not self intersect the reference polygon. The range of total number of vertices belonging to side visible to the sensor in any footprint is m where $1 < m \le n$. The inequality that $m > 1$ depicts that, if not occluded, at least one side or two vertices of the building are always visible to the side looking SAR sensor.

2) Compute the region of interest (ROI) by shifting/translating the identified vertices at a distance d towards the sensor.

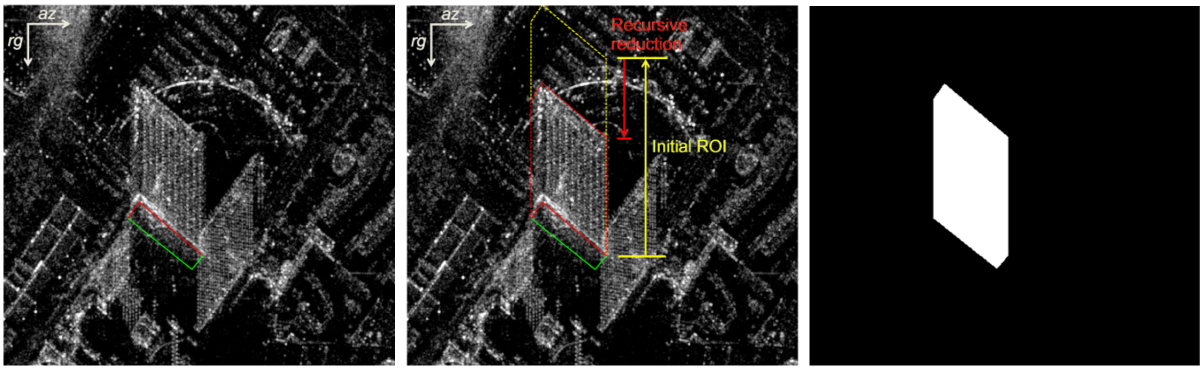
3) In ROI, mathematical morphology is applied to compute the number of background pixels, denoted as num. If I denote the ROI image, then num is computed as follows:

    - $I' = I \oplus S$ where $\oplus$ denotes the dilation operation and S is the structuring element used (line in our case).

- Apply nonparametric and unsupervised Ostu's thresholding technique on I' to obtain binary image B. The technique optimally determines a threshold by keeping the intra class variance to be minimum and is useful in reducing gray scale image to binary image by utilizing only the zeroth and first order cumulative moments of the gray level histogram [8].
- Compliment the image B to obtain B' and use connected components to find the largest component C in B'.
- If the number of pixels in the component C are N_C, then set $num = N_C$.

4) If *num* < *TH*, then create a mask using the polygonal ROI, otherwise set $d = d - 5$ and go to step 2. I.e., progressively shift the already shifted vertices 5m away from the sensor or towards the building footprint (see Figure 2).

Since the tallest building in the city of Las Vegas, the Stratosphere Tower, is around 350 m, *d* in step 1 is initialized to 385 i.e., maximum overlay of 385 pixels that a building could have in the SAR image of Las Vegas city, and is computed as 1.1×350 where 1.1m is the resolution of SAR sensor in the range direction.



(a) (b) (c)

Figure 2. Building mask extraction: (a) reference polygon (shown in red and green polylines) of the building of interest overlaid onto the SAR intensity map after geocoding. The side of the building facing the sensor is shown in red while the other side not visible to the sensor in green; (b) vertices connecting the red polyline are shifted towards the sensor as depicted by dotted yellow lines forming an initial ROI which is recursively reduced to the region shown by red dotted lines (TH = 25); (c) the final extracted building mask obtained upon termination of the recursive procedure.

3.2 Pixel Grouping

Following the extraction of building mask, pixels sharing similar height (ideally the same height) are then grouped together. This is done by translating the red polyline in Figure 2a and assigning nearest neighboring pixels to it, as illustrated in Figure 3. The color-coding already gives a rough idea about the relative height of façade w.r.t. ground.

4. Group Sparsity in TomoSAR

Instead of performing TomoSAR inversion pixel-wise in a single-snapshot fashion [5], we now exploit the multiple-snapshot data with the following group sparsity model.

4.1 Group Sparsity Model for TomoSAR – The M-SL1MMER Approach

For each pixel (snapshot) belonging to an iso-height group, by packing the data sampled at different spatial frequencies $\xi_n \forall n = \{1, \cdots, N\}$ as columns into the observation matrix $\mathbf{G} \in \mathbb{C}^{N \times M}$ and discretizing the elevation profile, subsequent linear equation system is obtained

$$\mathbf{G} = \mathbf{R}\,\mathbf{\Gamma} + \mathbf{E}, \tag{1}$$

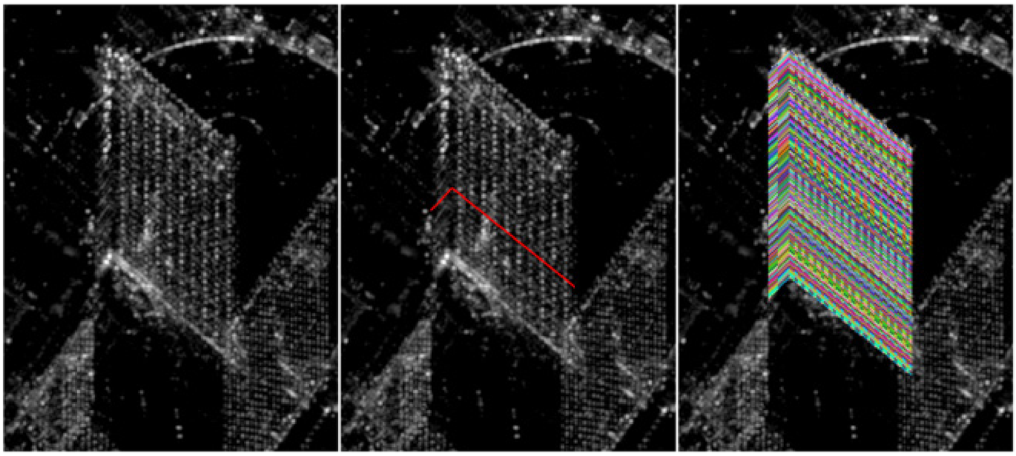
where $\mathbf{R} \in \mathbb{C}^{N \times L}$ is an overcomplete sensing matrix with L number of discretizations along the elevation axis, $\boldsymbol{\Gamma} \in \mathbb{C}^{L \times M}$ is the discrete reflectivity matrix of totally M number of snapshots, and $\mathbf{E} \in \mathbb{C}^{N \times M}$ the noise matrix. (1) is an underdetermined system with $N \ll L$ which has infinitely many solutions. However, if urban areas are concerned, each snapshot of $\boldsymbol{\Gamma}$ is characterized by only a small K number of dominant backscattering contributions (e.g., from ground-façade-roof interaction [7]), where K ranges typically from 0 to 3 [2]. This leads to the fact that the columns of $\boldsymbol{\Gamma}$ are K sparse. Since we expect all the iso-height pixels to share at least one common support, which corresponds to the contribution from façade, $\boldsymbol{\Gamma}$ is then row- or group-wise sparse.

To solve (1) with this prior, Malioutov et al. (2005) proposed the following method in [9]

$$\hat{\boldsymbol{\Gamma}} = \arg\min_{\boldsymbol{\Gamma}} \left\{ \frac{1}{2} \left\| \mathbf{G} - \mathbf{R}\boldsymbol{\Gamma} \right\|_{\mathrm{F}}^{2} + \lambda_K \left\| \boldsymbol{\Gamma} \right\|_{2,1} \right\}, \tag{2}$$

where $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm, λ_K is a hyperparameter balancing model error and the sparsity of $\boldsymbol{\Gamma}$, and $\|\cdot\|_{2,1}$ is $L_{2,1}$ mixed norm defined as the sum of the L_2 norm of each row and is known for promoting group sparsity. Note that similar models were utilized in [10] and [11], where either different polarimetric channels or adjacent pixels were exploited.

As done in the SL1MMER algorithm, the preliminary estimates $\hat{\boldsymbol{\Gamma}}$ will be refined by the follow-up model selection and parameter estimation.



(a) (b) (c)

Figure 3. Pixel grouping with (a) cropped intensity image; (b) one exemplary iso-height line; and (c) grouped iso-height pixels color-coded with group index.

4.2 Experiments using TanDEM-X data

The proposed M-SL1MMER is applied to TanDEM-X data mentioned in section 2. The results are compared to the ones obtained using SL1MMER. Figure 4 shows the reconstructed and color-coded elevation of the test building in Figure 2 and 3, overlaid with intensity. From top to bottom, 11 and 6 interferograms are used, respectively. From left to right, the separated superimposed scatterers, namely first layer with M-SL1MMER and SL1MMER, second layer with M-SL1MMER and SL1MMER, are illustrated respectively.

On the top of the test building, reflections from building roof and façade are overlaid. In Figure 4, dominating scattering (red) from roof can be seen in the first layer, whereas the corresponding parts of façade are visible in the second layer. Besides, parallelogram patterns in the second layer can be related to reflections within window frames. We do not expect many reflections from lower structures though, due to the large slope of the shell-like roof in front of the test building. It is evident that the multiple-snapshot model significantly outperforms the single-snapshot counterpart. In particular, when

N = 6, i.e. using extremely small number of images, the second layer estimated using SL1MMER is deteriorated by false alarms, while M-SL1MMER still achieves reasonable results.

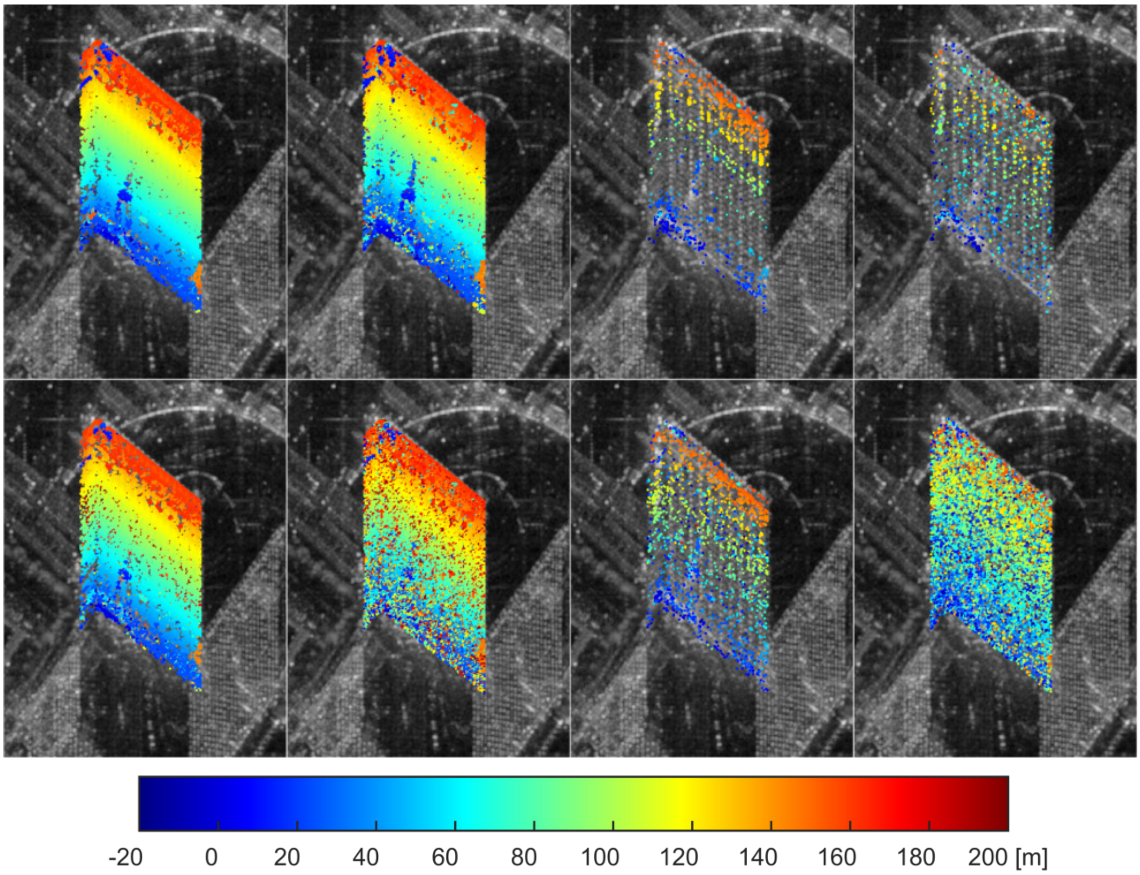


Figure 4. Reconstructed and color-coded elevation of the test building, overlaid with intensity. Top to bottom: N = 11 and 6, respectively; left to right: first layer with M-SL1MMER and SL1MMER, second layer with M-SL1MMER and SL1MMER, respectively.

5. Concluding Remarks

In this paper, a novel framework is proposed which can improve the estimation accuracy of TomoSAR inversion while reducing the required number of images. The core idea is the exploitation of group sparsity in iso-height SAR pixel groups that can be identified with the support of online available GIS data – 2D building footprints. Experiments on bistatic TanDEM-X data stacks demonstrate the great potential of the proposed approach. Further research will focus on the systematic performance analysis of the proposed framework, including the quality control and refinement of the GIS 2D footprints data, as well as super-resolution and estimation accuracy analysis of the M-SL1MMER algorithm.

References:

[1] S. Gernhardt, N. Adam, M. Eineder, and R. Bamler, "Potential of very high resolution SAR for persistent scatterer interferometry in urban areas," *Ann. GIS*, vol. 16, no. 2, pp. 103–111, Aug. 2010.

[2] X. X. Zhu and R. Bamler, "Very High Resolution Spaceborne SAR Tomography in Urban Environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 12, pp. 4296–4308, Dec. 2010.

[3] D. Reale, G. Fornaro, A. Pauciullo, X. Zhu, and R. Bamler, "Tomographic Imaging and Monitoring of Buildings With Very High Resolution SAR Data," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 661–665, Jul. 2011.

[4] X. X. Zhu and M. Shahzad, "Facade Reconstruction Using Multiview Spaceborne TomoSAR Point Clouds," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3541–3552, Jun. 2014.

[5] X. Zhu and R. Bamler, "Tomographic SAR Inversion by L1-Norm Regularization -- The Compressive Sensing Approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3839–3846, 2010.

[6] A. Budillon, A. Evangelista, and G. Schirinzi, "Three-Dimensional SAR Focusing From Multipass Signals Using Compressive Sampling," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 488–499, Jan. 2011.

[7] X. X. Zhu and R. Bamler, "Demonstration of Super-Resolution for Tomographic SAR Imaging in Urban Environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 8, pp. 3150–3157, Aug. 2012.

[8] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Trans. Syst. Man Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[9] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[10] E. Aguilera, M. Nannini, and A. Reigber, "Multi-signal compressed sensing for polarimetric SAR tomography," in *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*, 2011, pp. 1369–1372.

[11] M. Schmitt and U. Stilla, "Compressive Sensing Based Layover Separation in Airborne Single-Pass Multi-Baseline InSAR Data," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 2, pp. 313–317, Mar. 2013.